

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

JC968 U.S. PTO
10/075289
02/15/02



In re application of:

Sylvie LESMANNE ET AL.

Serial No.: To be assigned.

Filed: February 15, 2002

For: **COHERENCE CONTROLLER FOR A
MULTIPROCESSOR SYSTEM,
MODULE, AND MULTIPROCESSOR
SYSTEM WITH A MULTIMODULE
ARCHITECTURE INCORPORATING
SUCH A CONTROLLER**

: Examiner:

: Group Art Unit:

: Corresponding to:
FR 01/02089

: Filed February 15, 2001

McLean, Virginia

**CLAIM FOR BENEFIT OF FILING DATE
OF PRIOR FOREIGN APPLICATION**

Honorable Commissioner of Patents and Trademarks
Washington, DC 20231

Sir:.

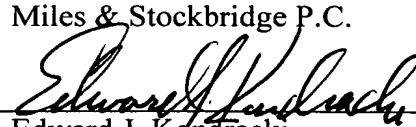
In the matter of the above-identified application, a claim is hereby made under the provisions of 35 U.S.C. 119 for the benefit of the filing date of the corresponding French application No 01/02089 filed February 15, 2001, which is referred to in the Declaration of the present case.

A certified copy of said French application is attached hereto.

Respectfully submitted,

Miles & Stockbridge P.C.

By:


Edward J. Kondracki
Registration No. 20,604

Date: February 15, 2002

Miles & Stockbridge, P.C.
1751 Pinnacle Drive, Suite 500
McLean, Virginia 22102-3833
Tel.: (703) 903-9000

THIS PAGE BLANK (USPTO)



JC868 U.S. PRO
10/075289



BREVET D'INVENTION

CERTIFICAT D'UTILITÉ - CERTIFICAT D'ADDITION

COPIE OFFICIELLE

Le Directeur général de l'Institut national de la propriété industrielle certifie que le document ci-annexé est la copie certifiée conforme d'une demande de titre de propriété industrielle déposée à l'Institut.

21 JAN. 2002

Fait à Paris, le

Pour le Directeur général de l'Institut
national de la propriété industrielle
Le Chef du Département des brevets

Martine PLANCHE

**INSTITUT
NATIONAL DE
LA PROPRIETE
INDUSTRIELLE**

SIEGE
26 bis, rue de Saint Petersbourg
75800 PARIS cedex 08
Téléphone : 33 (1) 53 04 53 04
Télécopie : 33 (1) 42 93 59 30
www.inpi.fr

THIS PAGE BLANK (USPTO)

BREVET D'INVENTION

CERTIFICAT D'UTILITÉ

Code de la propriété intellectuelle - Livre VI

REQUÊTE EN DÉLIVRANCE 1/2

Cet imprimé est à remplir lisiblement à l'encre noire

DB 540 W /190600

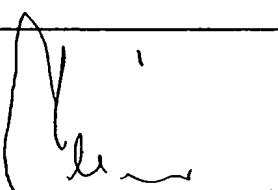
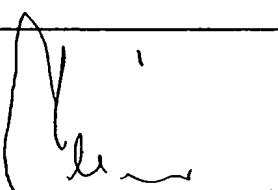
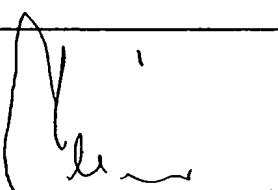
REMISE DES PIÈCES DATE 15 FEV 2001 LIEU 75 INPI PARIS N° D'ENREGISTREMENT NATIONAL ATTRIBUÉ PAR L'INPI 0102089 DATE DE DÉPÔT ATTRIBUÉE PAR L'INPI 15 FEV. 2001		Réserve à l'INPI 1 NOM ET ADRESSE DU DEMANDEUR OU DU MANDATAIRE À QUI LA CORRESPONDANCE DOIT ÊTRE ADRESSÉE BULL S.A. Monsieur DENIS Hervé 68, route de Versailles PC : 58D20 78434 LOUVECIENNES Cedex	
Vos références pour ce dossier <i>(facultatif)</i> FR 3933 HD			
Confirmation d'un dépôt par télécopie <input type="checkbox"/> N° attribué par l'INPI à la télécopie			
2 NATURE DE LA DEMANDE Demande de brevet <input checked="" type="checkbox"/> Demande de certificat d'utilité <input type="checkbox"/>		Cochez l'une des 4 cases suivantes <input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> N° _____ Date _____ / _____ / _____ N° _____ Date _____ / _____ / _____ <input type="checkbox"/> N° _____ Date _____ / _____ / _____	
3 TITRE DE L'INVENTION (200 caractères ou espaces maximum) Contrôleur de cohérence pour ensemble multiprocesseur, module et ensemble multiprocesseur à architecture multimodule intégrant un tel contrôleur.			
4 DÉCLARATION DE PRIORITÉ OU REQUÊTE DU BÉNÉFICE DE LA DATE DE DÉPÔT D'UNE DEMANDE ANTÉRIEURE FRANÇAISE		Pays ou organisation Date _____ / _____ / _____ N° Pays ou organisation Date _____ / _____ / _____ N° Pays ou organisation Date _____ / _____ / _____ N° <input type="checkbox"/> S'il y a d'autres priorités, cochez la case et utilisez l'imprimé « Suite »	
5 DEMANDEUR		<input type="checkbox"/> S'il y a d'autres demandeurs, cochez la case et utilisez l'imprimé « Suite »	
Nom ou dénomination sociale Prénoms Forme juridique N° SIREN Code APE-NAF Adresse _____ Rue _____ Pays _____ Nationalité _____ N° de téléphone <i>(facultatif)</i> N° de télécopie <i>(facultatif)</i> Adresse électronique <i>(facultatif)</i>		BULL S.A. Société Anonyme 6 4 2 0 5 8 7 3 9 3 0 0 C 68, route de Versailles 78430 LOUVECIENNES France Française 01.39.66.61.80 01.39.66.61.73 hervé.denis@bull.net	

**BREVET D'INVENTION
CERTIFICAT D'UTILITÉ**

REQUÊTE EN DÉLIVRANCE 2/2

Réservé à l'INPI	
REMISE DES PIÈCES	
DATE	15 FEV 2001
LIEU	75 INPI PARIS
N° D'ENREGISTREMENT	
NATIONAL ATTRIBUÉ PAR L'INPI	0102089

DB 540 W /190600

Vos références pour ce dossier : (facultatif)		FR 3933 HD																		
6 MANDATAIRE <table border="1" style="width: 100%;"> <tr> <td style="width: 50%;">Nom</td> <td>DENIS</td> </tr> <tr> <td>Prénom</td> <td>Hervé</td> </tr> <tr> <td>Cabinet ou Société</td> <td>BULL S.A.</td> </tr> <tr> <td>N ° de pouvoir permanent et/ou de lien contractuel</td> <td>PG 4972</td> </tr> <tr> <td>Adresse</td> <td>68, route de Versailles</td> </tr> <tr> <td>Code postal et ville</td> <td>78430 LOUVECIENNES</td> </tr> <tr> <td>N° de téléphone (facultatif)</td> <td>01.39.66.61.80</td> </tr> <tr> <td>N° de télécopie (facultatif)</td> <td>01.39.66.61.73</td> </tr> <tr> <td>Adresse électronique (facultatif)</td> <td>hervé.denis@bull.net</td> </tr> </table>			Nom	DENIS	Prénom	Hervé	Cabinet ou Société	BULL S.A.	N ° de pouvoir permanent et/ou de lien contractuel	PG 4972	Adresse	68, route de Versailles	Code postal et ville	78430 LOUVECIENNES	N° de téléphone (facultatif)	01.39.66.61.80	N° de télécopie (facultatif)	01.39.66.61.73	Adresse électronique (facultatif)	hervé.denis@bull.net
Nom	DENIS																			
Prénom	Hervé																			
Cabinet ou Société	BULL S.A.																			
N ° de pouvoir permanent et/ou de lien contractuel	PG 4972																			
Adresse	68, route de Versailles																			
Code postal et ville	78430 LOUVECIENNES																			
N° de téléphone (facultatif)	01.39.66.61.80																			
N° de télécopie (facultatif)	01.39.66.61.73																			
Adresse électronique (facultatif)	hervé.denis@bull.net																			
7 INVENTEUR (S) <table border="1" style="width: 100%;"> <tr> <td style="width: 50%;">Les inventeurs sont les demandeurs</td> <td><input type="checkbox"/> Oui <input checked="" type="checkbox"/> Non Dans ce cas fournir une désignation d'inventeur(s) séparée</td> </tr> </table>			Les inventeurs sont les demandeurs	<input type="checkbox"/> Oui <input checked="" type="checkbox"/> Non Dans ce cas fournir une désignation d'inventeur(s) séparée																
Les inventeurs sont les demandeurs	<input type="checkbox"/> Oui <input checked="" type="checkbox"/> Non Dans ce cas fournir une désignation d'inventeur(s) séparée																			
8 RAPPORT DE RECHERCHE <table border="1" style="width: 100%;"> <tr> <td style="width: 50%;">Établissement immédiat ou établissement différé</td> <td><input checked="" type="checkbox"/> <input type="checkbox"/></td> </tr> <tr> <td>Paiement échelonné de la redevance</td> <td>Paiement en deux versements, uniquement pour les personnes physiques <input type="checkbox"/> Oui <input type="checkbox"/> Non</td> </tr> </table>			Établissement immédiat ou établissement différé	<input checked="" type="checkbox"/> <input type="checkbox"/>	Paiement échelonné de la redevance	Paiement en deux versements, uniquement pour les personnes physiques <input type="checkbox"/> Oui <input type="checkbox"/> Non														
Établissement immédiat ou établissement différé	<input checked="" type="checkbox"/> <input type="checkbox"/>																			
Paiement échelonné de la redevance	Paiement en deux versements, uniquement pour les personnes physiques <input type="checkbox"/> Oui <input type="checkbox"/> Non																			
9 RÉDUCTION DU TAUX DES REDEVANCES <table border="1" style="width: 100%;"> <tr> <td style="width: 50%;">Si vous avez utilisé l'imprimé «Suite», indiquez le nombre de pages jointes</td> <td>Uniquement pour les personnes physiques <input type="checkbox"/> Requise pour la première fois pour cette invention (<i>joindre un avis de non-imposition</i>) <input type="checkbox"/> Requise antérieurement à ce dépôt (<i>joindre une copie de la décision d'admission pour cette invention ou indiquer sa référence</i>)</td> </tr> </table>			Si vous avez utilisé l'imprimé «Suite», indiquez le nombre de pages jointes	Uniquement pour les personnes physiques <input type="checkbox"/> Requise pour la première fois pour cette invention (<i>joindre un avis de non-imposition</i>) <input type="checkbox"/> Requise antérieurement à ce dépôt (<i>joindre une copie de la décision d'admission pour cette invention ou indiquer sa référence</i>)																
Si vous avez utilisé l'imprimé «Suite», indiquez le nombre de pages jointes	Uniquement pour les personnes physiques <input type="checkbox"/> Requise pour la première fois pour cette invention (<i>joindre un avis de non-imposition</i>) <input type="checkbox"/> Requise antérieurement à ce dépôt (<i>joindre une copie de la décision d'admission pour cette invention ou indiquer sa référence</i>)																			
10 SIGNATURE DU DEMANDEUR OU DU MANDATAIRE (Nom et qualité du signataire) <table border="1" style="width: 100%;"> <tr> <td style="width: 50%;"></td> <td>VISA DE LA PRÉFECTURE OU DE L'INPI</td> </tr> <tr> <td>Hervé DENIS (Mandataire Bull S.A)</td> <td>M. MARTIN</td> </tr> </table>				VISA DE LA PRÉFECTURE OU DE L'INPI	Hervé DENIS (Mandataire Bull S.A)	M. MARTIN														
	VISA DE LA PRÉFECTURE OU DE L'INPI																			
Hervé DENIS (Mandataire Bull S.A)	M. MARTIN																			

CONTROLEUR DE COHERENCE POUR ENSEMBLE MULTIPROCESSEUR,
MODULE ET ENSEMBLE MULTIPROCESSEUR A ARCHITECTURE MULTI-
MODULE INTEGRANT UN TEL CONTROLEUR

5 La présente invention concerne la réalisation d'ensembles multiprocesseurs symétriques de grande taille par assemblage de multiprocesseurs de base plus petits et comportant chacun en général de un à quatre microprocesseurs élémentaires (μ P) associés chacun à une
10 mémoire cache, une mémoire principale (MEM) et un circuit d'entrée et sortie (E/S) convenablement reliés entre eux par un réseau de bus approprié. L'ensemble multiprocesseur étant géré par un système d'exploitation OS commun (également appelé en anglais **Operating**
15 **System**). En particulier l'invention concerne des contrôleurs de cohérence intégrés aux ensembles multiprocesseurs et destinés à assurer la cohérence mémoire de ces derniers notamment entre mémoires principales et caches, étant spécifié qu'une procédure
20 d'accès à la mémoire est considérée comme « cohérente » si la valeur en retour à une instruction de lecture est toujours la valeur écrite par la dernière instruction de mémorisation. Dans la pratique on rencontre des incohérences sur mémoires caches d'une part dans des
25 procédures d'Entrées/Sorties et d'autre part dans des situations où l'écriture immédiate dans la mémoire d'un multiprocesseur est autorisée sans avoir attendu et vérifié que tous les caches susceptibles d'avoir une copie de la mémoire aient été modifiés.

30 On connaît des multiprocesseurs réalisés selon le schéma de principe illustré à la figure 1 et à titre d'exemple non limitatif principalement constitués de quatre multiprocesseurs de base 10-13 MP0, MP1, MP2 et MP3, à deux microprocesseurs 40 et 40', respectivement reliés à un contrôleur de cohérence 14 SW (**Switch**) par des liaisons bi-point haut débit 20-23 gérées par quatre

unités de gestion de port local 30-33 PU0, PU1, PU2 et PU3. Le contrôleur 14 connaît la répartition de la mémoire et les copies de lignes ou blocs mémoire entre la mémoire principale MEM 44 et les mémoires caches 42, 5 42' des processeurs et comprend, outre une ou plusieurs tables de routage et une table de fenêtre de collision (non représentées), un répertoire de filtrage de caches 34 SF (également appelé en anglais **Snoop Filter**) qui garde la trace des copies de portions de mémoire (lignes 10 ou blocs) présentes dans les caches des multiprocesseurs. Pour la suite de l'exposé et par convention les termes « lignes » ou « blocs » seront, sauf indication contraire, utilisées indifféremment pour désigner l'un ou l'autre terme. Par ailleurs, le terme 15 « mémoire » utilisé seul concerne la ou les mémoires principales associées aux multiprocesseurs.

Le répertoire de filtrage des caches 34, géré par l'unité de gestion ILU 15, est capable de transmettre 20 des requêtes d'accès cohérent à un bloc mémoire (en vue d'une opération ultérieure telle que Lecture, Ecriture, Effacement, etc ..) soit à la mémoire principale concernée soit au(x) multiprocesseur(s) ayant une copie du bloc recherché dans leurs caches, après vérification 25 de l'état mémoire du bloc concerné, afin de préserver la cohérence mémoire de l'ensemble. Pour ce faire le répertoire de filtrage des caches 34 comporte l'adresse 35 de chaque bloc répertorié associée à un vecteur de présence 36 de 4 bits (où 4 représente le nombre « n » 30 de multiprocesseurs de base 10-13) et à un bit Ex d'état mémoire Exclusif 37.

Dans la pratique le bit MPO du vecteur de présence 36 est porté à 1 lorsque le multiprocesseur de base MPO 35 correspondant (le multiprocesseur 10) comporte effectivement dans une de ses mémoires caches une copie d'une ligne ou d'un bloc de la mémoire 44.

Le bit d'état Exclusif Ex 37 appartient au Protocole de cohérence dit Protocole MESI qui d'une façon générale décrit les quatre états mémoire suivants :

- 5 Modifié (**Modified**) : selon lequel le bloc (ou la ligne) dans le cache a été modifié(e) par rapport au contenu de la mémoire (la donnée dans le cache est valide mais de la position mémoire correspondante est périmee).
- 10 Exclusif (**Exclusive**) : selon lequel le bloc dans le cache contient l'unique copie à l'identique des données de la mémoire aux mêmes adresses.
- 15 Partagé (**Shared**) : selon lequel le bloc dans le cache contient des données identiques à celles de la mémoire aux mêmes adresses (au moins un autre cache peut disposer des mêmes données).
- 20 Invalide (**Invalid**) : selon lequel les données dans le bloc sont invalides et ne peuvent être utilisées.

- Dans la pratique, on utilise pour les multiprocesseurs illustrés à la figure 1 et la figure 2 un protocole MESI partiel selon lequel les états « Modifié » et « Exclusif » ne sont pas distingués :
 - si un seul bit MPi = 1 et si le bit Ex = 1, alors l'état mémoire du bloc (ou de la ligne) sera Modifié ou Exclusif ;
 - si un ou plusieurs bits MPi = 1 et si le bit Ex = 0, alors l'état mémoire du bloc sera Partagé ;
 - si tous les bits MPi = 0, alors l'état mémoire sera Invalide.

- 30 Le répertoire de filtrage des caches 34 intègre un protocole de recherche et de surveillance doté d'une logique dite de « **snooping** ». Ainsi lors d'une requête d'accès en mémoire par un processeur le répertoire de filtrage de caches 34 procède à l'examen des mémoires caches qu'il gère. Lors de cette vérification le trafic s'effectue au travers des ports 24-27 des liaisons au

haut débit bi-point 20-23 sans interférer avec les accès entre le processeur 40 et sa mémoire cache 42. Le répertoire de filtrage des caches est alors capable de traiter toutes les requêtes d'accès mémoire cohérents.

5

L'architecture de multiprocesseur connue et brièvement décrite ci-avant n'est toutefois pas adaptée à des applications de grands serveurs multiprocesseurs symétriques de grandes tailles comportant plus de 16 processeurs.

10 En effet le nombre de multiprocesseurs de base pouvant être connectés à un contrôleur de cohérence (en pratique réalisé par un circuit intégré du type ASIC) est limité 15 en pratique par :

- le nombre d'entrées /sorties du contrôleur qui selon les techniques de fabrication actuelles n'accepte qu'un nombre limité de liaisons bi-point (étant fait remarquer que ces liaisons bi-point sont nécessaires grâce à leur 20 capacité de haut débit pour éviter les problèmes de latence ou d'attente lors des traitements des requêtes d'accès mémoire).
- la taille du contrôleur de cohérence qui contient le 25 répertoire de filtrage des caches (la taille du répertoire de filtrage de caches doit être supérieure à la somme des tailles des répertoires des caches intégrés dans les multiprocesseurs de base).
- la bande passante d'accès au répertoire de filtrage des caches ou débit maximal en Mbits/sec obtenu en 30 pratique par les liaisons bi-point constitue un goulot d'étranglement pour un serveur multiprocesseur de grande taille, le répertoire de filtrage de caches devant être consulté pour tous les accès cohérents des multiprocesseurs de base.

35

La présente invention a pour but de proposer un contrôleur de cohérence susceptible notamment de

remédier aux inconvénients présentés ci-dessus ou d'en atténuer sensiblement les effets. L'invention a également pour objet de proposer des ensembles multiprocesseurs de grande taille à architecture multi-module, notamment des serveurs multiprocesseurs symétriques, à performances améliorées.

A cette fin l'invention propose un contrôleur de cohérence adapté pour être connecté à une pluralité de processeurs équipés d'une mémoire cache et à au moins une mémoire principale locale pour définir un module local de multiprocesseurs de base, ledit contrôleur de cohérence comportant un répertoire de filtrage de caches comportant un premier répertoire de filtrage SF destiné à assurer la cohérence entre la mémoire principale locale et les mémoires caches du module local, caractérisé en ce qu'il comporte de plus un port externe adapté pour être connecté à au moins un module de multiprocesseurs externe identique ou compatible avec ledit module local, le répertoire de filtrage de caches comportant un répertoire de filtrage complémentaire ED pour garder trace des coordonnées, notamment les adresses, des lignes ou blocs de la mémoire principale locale copiés du module local vers un module externe et assurer la cohérence entre la mémoire principale locale et les mémoires caches du module local et des modules externes.

Ainsi l'extension ED du répertoire de filtrage de caches est gérée comme le répertoire de filtrage de caches SF et permet de savoir s'il existe des copies de la mémoire du module local hors de ce module et de ne propager les requêtes d'origine locale qu'à bon escient vers les autres modules ou modules externes.

35

Cette solution trouve sa pleine efficacité dans les systèmes d'exploitation actuels qui commencent à gérer

les affinités entre processus en cours de traitement et la mémoire qu'ils utilisent (avec regroupement automatique entre mémoires et multiprocesseurs concernés). Dans ce cas la taille du répertoire ED nécessaire peut être inférieure à celle du répertoire SF et la bande passante de la liaison inter-modules peut être inférieure au double d'une liaison intra-module.

10 Selon un mode de réalisation préférentiel du contrôleur de cohérence selon l'invention, le contrôleur de cohérence comporte un vecteur de présence à « n » bits, où n est le nombre de multiprocesseurs de base dans un module (vecteur de présence locale), une extension du vecteur de présence de « N-1 » bits, où N-1 est le nombre total de modules externes connectés à la liaison externe (extension de présence distante), et un bit de statut Exclusif. Ainsi seules les lignes ou blocs de la mémoire locale peuvent avoir un vecteur de présence non nul dans le répertoire de filtrage de caches ED.

20 Cette caractéristique est également très intéressante car elle permet sans difficultés particulières de gérer sensiblement de la même façon les liaisons inter-modules et les liaisons intra-module, le protocole de gestion de 25 contrôleur de cohérence étant étendu pour prendre en compte la notion de mémoire locale ou de mémoire distante dans les modules externes.

30 Avantageusement le contrôleur de cohérence comporte n unités de gestion des ports locaux PU connectés au n multiprocesseurs de base du module local et une unité de gestion XPU du port externe et une unité commune de gestion ILU des répertoires de filtrage SF et ED. De même l'unité de gestion XPU du port externe et les 35 unités de gestion PU des ports locaux sont compatibles entre elles et utilisent des protocoles voisins largement communs.

5 L'invention concerne également un module de multiprocesseurs comportant une pluralité de processeurs équipés d'une mémoire cache et au moins une mémoire principale connectés à un contrôleur de cohérence défini ci-avant dans ses différentes versions.

10 L'invention concerne également un ensemble multiprocesseur à architecture multi-module comportant au moins deux modules de multiprocesseurs selon l'invention défini ci-dessus connectés entre eux directement ou indirectement par les liaisons externes des répertoires de filtrage des caçhes de leurs contrôleurs de cohérence.

15 20 25 Avantageusement les liaisons externes de l'ensemble multiprocesseur à architecture multi-module sont connectées entre elles travers d'un dispositif de commutation ou routeur. Tout aussi avantageusement le dispositif de commutation ou routeur comporte des moyens de gestion et/ou de filtrage des données et/ou requêtes en transit.

30 35 L'invention concerne également un serveur multiprocesseur architecture multi-module symétrique de grande taille, comportant « N » modules de multiprocesseurs identiques ou compatibles entre eux, chaque module comportant une pluralité de « n » multiprocesseurs de base équipés d'au moins une mémoire cache et d'au moins une mémoire principale locale et connectés à un contrôleur de cohérence local comportant un répertoire de filtrage local de caches SF destiné à assurer la cohérence locale entre la mémoire principale locale et les mémoires caches du module ci-après appelé module local, chaque contrôleur de cohérence local étant connecté par une liaison externe bi-point, éventuellement par l'intermédiaire d'un dispositif commutateur ou routeur, à au moins un module de

multiprocesseurs externe audit module local, le contrôleur de cohérence comportant un répertoire de filtrage de caches complémentaire ED pour garder trace des coordonnées, notamment les adresses, des lignes ou 5 blocs mémoire copiés du module local vers un module externe et assurer la cohérence entre la mémoire principale locale et les mémoire caches du module local et des modules externes.

10 Selon un mode de réalisation préférentiel du serveur multiprocesseur à architecture multi-module selon l'invention, chaque contrôleur de cohérence comporte un vecteur de présence à « n » bits destiné à indiquer la présence ou l'absence d'une copie d'un bloc ou ligne 15 mémoire dans les mémoires caches des multiprocesseurs de base locaux (vecteur de présence locale), une extension du vecteur de présence de « N-1 » bits destinée à indiquer la présence ou l'absence d'une copie d'un bloc ou d'une ligne mémoire dans les mémoires caches des 20 multiprocesseurs des modules externes (extension de présence distante) et un bit Ex de statut Exclusif.

Avantageusement le dispositif de commutation ou routeur comporte des moyens de gestion et/ou de filtrage des 25 données et/ou requêtes en transit.

D'autres buts, avantages et caractéristiques de l'invention apparaîtront à la lecture de la description qui va suivre d'un mode de réalisation d'un contrôleur 30 de cohérence et d'un serveur multiprocesseur à architecture multi-module selon l'invention donnés à titre d'exemple non limitatif en référence aux dessins ci-annexés dans lesquels:

- la figure 1 montre une représentation schématique 35 d'un serveur multiprocesseur selon un art antérieur connu et présenté dans le préambule du présent exposé; et

- la figure 2 montre une représentation schématique d'un serveur multiprocesseur à architecture multi-module selon l'invention avec un contrôleur de cohérence à fonction étendue selon l'invention.

5

L'ensemble ou serveur multiprocesseur à architecture multi-module illustré schématiquement à la figure 2 est principalement constitué de quatre ($N = 4$) modules 50-53 (Mod 0 à Mod 3) identiques ou compatibles entre eux et 10 convenablement reliés entre eux au travers d'un dispositif de commutation ou routeur 54 par des liaisons haut débit bi-point respectivement 55 à 58. Par commodité seul le Mod 0 50 est illustré en détails à la figure 2.

15

A titre d'exemple non limitatif et par commodité de description, chaque module 50-53 est constitué de $n = 4$ ensembles multiprocesseurs de base 60-63 MPO-MP3 respectivement reliés à un contrôleur de cohérence 64 20 SW (Switch) par des liaisons bi-point haut débit 70-73 gérées par quatre unités de gestion PU0, PU1, PU2 et PU3 80-83 de port local 90-93. Toujours à titre d'exemple non limitatif chaque multiprocesseur de base MPO-MP3 60-63 est identique au multiprocesseur 10 déjà décrit en 25 référence à la figure 1 et comporte deux processeurs 40, 40' avec leur mémoires caches 42, 42', au moins une mémoire principale commune, une unité d'entrée/sortie associés par un réseau de bus commun. D'une façon générale la structure et le mode opératoire des modules 30 50-53 sont voisins du serveur à multiprocesseur de la figure 1 et seront pas repris en détails tout du moins en ce qui concerne les points communs aux deux serveurs à multiprocesseurs. En particulier le serveur multiprocesseur à architecture multi-module de 35 l'invention est également géré par un système d'exploitation du type OS commun à tous les modules.

Pour assurer la cohérence locale des accès mémoire au niveau de chaque module, le contrôleur de cohérence 64 de chaque module (par exemple le module 50) comporte un répertoire de filtrage étendu de caches SF/ED 84 auquel 5 est affecté une double fonction :

- la fonction classique de « Snoop Filter » (SF) assurée localement au sein du module intégrant le contrôleur de cohérence concerné qui garde la trace des copies de portions de mémoire (lignes ou blocs) présentes dans les 10 caches des huit processeurs présents dans le même module (en l'espèce le module 50) et présentée ci-avant en référence à la figure 1 ;
- la fonction étendue de répertoire externe (ED) qui garde la trace des lignes ou blocs de mémoire locale 15 (c.à.d. appartenant au module 50) exportées vers les autres modules 51, 52 et 53.

Pour ce faire le répertoire de filtrage des caches 84, contrôlé par l'unité de gestion 65, comporte l'adresse 20 85 de chaque bloc répertorié associée à un vecteur de présence locale 86 de 4 bits (où 4 représente le nombre « n » de multiprocesseurs de base 60-63) et à un bit Ex d'état mémoire Exclusif 87 dont les caractéristiques et 25 la fonction ont déjà été présentées en référence au serveur de la figure 1. Dans la pratique le bit MPO du vecteur de présence 86 est porté à 1 lorsque le multiprocesseur de base MPO correspondant (le multiprocesseur 60) comporte effectivement dans une de ses mémoires caches une copie d'une ligne ou d'un bloc 30 de la mémoire principale intégrée à ce multiprocesseur MPO. De plus il est prévu une extension de présence distante 88 sur 3 bits du vecteur de présence (ou 3 représente le nombre N-1, avec N = 4 égal au nombre de modules du serveur multiprocesseur), le bit Mod1 de 35 l'extension 88 étant porté à 1 lorsque le module 51 (le module Mod 1) comporte effectivement dans une de ses mémoires caches une copie d'une ligne ou d'un bloc

mémoire appartenant au module 50 Mod 0. Dans la pratique le répertoire de filtrage de caches 84 SF/ED est constitué par la fusion des répertoires de filtrage SF et ED, étant fait remarquer que seules les lignes de la 5 mémoire locale peuvent avoir une extension de vecteur de présence non nulle dans le répertoire ED.

Pour terminer, le contrôleur de cohérence 64 comporte une unité de gestion XPU 89 gérant le port externe 99 10 convenablement relié à la liaison bi-point 55 connectée au routeur 54. Dans la pratique les unités PU0-PU3 60-63 et XPU 89 utilisent des protocoles très voisins, notamment des protocole de communication, et ont sensiblement le même comportement :

15 - Pour toute requête d'accès cohérent venant d'un port local ou externe, l'unité (X)PU concernée transmet la requête à ILU 65 qui :

16 _ renvoie au (X)PU émetteur l'état du répertoire de filtrage des caches,
20 _ transmet la requête aux unités ayant une copie si nécessaire,
25 _ ouvre dans l'ILU une fenêtre de collision si nécessaire (pour assurer un traitement en série et exhaustif des requêtes en cas de collision de requêtes associées à une même adresse mémoire).

30 - Pour toute requête émise par ILU, l'unité (X)PU concernée transmet la requête sur le port associé et transmet au destinataire toutes les réponses reçues du port.

35 - Les unités (X)PU gèrent les réponses attendues pour une requête cohérente et une fois toutes les réponses arrivées, ces unités (X)PU ferment la fenêtre de collision et demandent la mise à jour du répertoire de filtrage de caches avec les bons bits de présence et de statut. Un module qui émet des requêtes vers extérieur reçoit toujours une réponse pour fermer sa fenêtre de collision et mettre à jour son répertoire SF/ED.

Par ailleurs un échec (en anglais « **miss** ») pour la recherche d'une adresse locale dans le répertoire SF/ED entraîne un routage vers l'unité de port locale PU du module résidant (« **home** ») de l'adresse recherchée. De même un échec (« **miss** ») pour la recherche d'une adresse distante dans le répertoire SF/ED entraîne un routage vers l'unité de port externe XPU.

On notera que la fenêtre de collision principale est gérée dans le module résidant « **home** » avec une fenêtre de collision annexe gérée dans le module émetteur pour qu'un module n'émette qu'une seule requête sur la même adresse (réessais compris) et une fenêtre de collision annexe gérée dans le module cible pour que le répertoire SF/ED ne reçoive qu'une seule requête sur la même adresse.

Parmi les différences rencontrées entre les unités PU et XPU, on notera également que les requêtes/réponses envoyées sur le port externe sont accompagnées d'un masque porteur d'informations complémentaires désignant le ou les modules destinataires parmi les N-1 autres modules. Enfin sur une ligne distante un accès « **miss** » dans SF/ED si émis par PU est transmis sur le port externe et si émis par XPU recevra pour réponse le message « pas de copie locale ».

Ainsi le contrôleur de cohérence selon l'invention à port externe et à répertoire de filtrage de caches à vecteur de présence étendu et son implémentation dans un ensemble multiprocesseurs à architecture multi-module permet du gain substantiel dans la taille des répertoires de filtrage de caches et au niveau de la bande passante par rapport à une simple extrapolation du multiprocesseur de l'art antérieur présenté ci-avant.

L'invention n'est pas limitée à un ensemble multiprocesseurs à architecture multi-module à 32 processeurs ici décrit à titre d'exemple non limitatif mais concerne également des ensembles ou serveurs 5 multiprocesseurs à 64 processeurs ou plus. De même sans sortir du cadre de l'invention le routeur 54 décrit comme un dispositif de commutation de base comporte des moyens de gestion et/ou de filtrage des données et/ou requêtes en transit.

REVENDICATIONS:

1. Contrôleur de cohérence (64) adapté pour être connecté à une pluralité de processeurs (40, 40') équipés d'une mémoire cache (42, 42') et à au moins une mémoire principale locale (44) pour définir un module local (50) de multiprocesseurs de base (60), ledit contrôleur de cohérence (64) comportant un répertoire de filtrage de caches (84) comportant un premier répertoire de filtrage 10 SF destiné à assurer la cohérence entre la mémoire principale locale (44) et les mémoires caches (42, 42') du module local, caractérisé en ce qu'il comporte de plus un port externe (99) adapté pour être connecté à au moins un module de multiprocesseurs externe (51, 52, 53) 15 identique ou compatible avec ledit module local (50), le répertoire de filtrage de caches (84) comportant un répertoire de filtrage complémentaire ED pour garder trace des coordonnées, notamment les adresses, des lignes ou blocs de la mémoire principale locale (44) 20 copiés du module local (50) vers un module externe (51, 52, 53) et assurer la cohérence entre la mémoire principale locale (44) et les mémoires caches (42, 42') du module local (50) et des modules externes (51, 52, 53).
25
2. Contrôleur de cohérence (64) selon la revendication 1 caractérisé en ce qu'il comporte outre un vecteur de présence (86) à « n » bits, où n est le nombre de multiprocesseurs de base dans un module, une extension 30 (88) de « N-1 » bits du vecteur de présence où N-1 est le nombre total de modules externes (51, 52, 53) connectés au port externe (99), et un bit de statut Exclusif (87).
3. Contrôleur de cohérence (64) selon la revendication 2 35 caractérisé en ce que le port externe (99) est connecté directement ou indirectement aux modules externes (51,

52, 53) par l'intermédiaire d'une liaison externe bi-point (55).

4. Contrôleur de cohérence (64) selon la revendication 2
5 caractérisé en ce qu'il comporte « n » unités de gestion
PU (80-83) des ports locaux (90-93) connectées aux n
multiprocesseurs de base (60-63) du module local (50),
une unité de gestion XPU (89) du port externe (99) et
une unité commune de gestion ILU des répertoires de
10 filtrage SF/ED (84)..

5. Contrôleur de cohérence (64) selon la revendication 4
caractérisé en ce que l'unité de gestion XPU (89) du
port externe et les unités de gestion PU (80-83) des
15 ports locaux sont compatibles entre elles et utilisent
des protocoles voisins largement communs.

6. Module (50) de multiprocesseurs caractérisé en ce
qu'il comporte une pluralité de multiprocesseurs (60-63)
20 équipés d'au moins une mémoire cache (42, 42') et au
moins une mémoire principale (44) et connectés à un
contrôleur de cohérence (64) selon l'une des
revendications 1 à 5.

25 7. Ensemble multiprocesseur à architecture multi-module
caractérisé en ce qu'il comporte au moins deux modules
de multiprocesseurs (50-53) selon la revendication 6
connectés entre eux directement ou indirectement par les
ports externes (99) de leurs contrôleurs de cohérence
30 (64).

8. Ensemble multiprocesseur selon la revendication 7,
caractérisé en ce que lesdits ports externes (99) sont
connectés entre eux au travers d'un dispositif de
35 commutation ou routeur (54).

9. Ensemble multiprocesseur selon la revendication 8, caractérisé en ce que le dispositif de commutation ou routeur (54) comporte des moyens de gestion et/ou de filtrage des données et/ou requêtes en transit.

5

10. Serveur multiprocesseur à architecture multi-module symétrique de grande taille, caractérisé en ce qu'il comporte « N » modules de multiprocesseurs (50-53) identiques ou compatibles entre eux, chaque module 10 comportant une pluralité de « n » multiprocesseurs de base (60-63) équipés d'au moins une mémoire cache (42) et d'au moins une mémoire principale locale (44) et connectés à un contrôleur de cohérence local (64) comportant un répertoire de filtrage local de caches SF 15 destiné à assurer la cohérence locale entre la mémoire principale locale et les mémoires caches du module ci-après appelé module local, chaque contrôleur de cohérence local (64) étant connectée par une liaison externe bi-point (55), éventuellement par 20 l'intermédiaire d'un dispositif commutateur ou routeur (54), à au moins un module de multiprocesseurs (51, 52, 53) externe audit module local, le contrôleur de cohérence (64) comportant un répertoire de filtrage de caches complémentaire ED pour garder trace des 25 coordonnées, notamment les adresses, des lignes ou blocs mémoire copiés du module local vers un module externe et assurer la cohérence entre la mémoire principale locale (44) et les mémoire caches (42, 42') du module local (50) et des modules externes (51, 52, 53).

30

11. Serveur multiprocesseur à architecture multi-module selon la revendication 10 caractérisé en ce que chaque contrôleur de cohérence (64) comporte un vecteur de présence (86) à « n » bits destiné à indiquer la présence ou l'absence d'une copie d'un bloc ou ligne mémoire dans les mémoires caches des multiprocesseurs de base locaux, une extension (88) du vecteur de présence

de « N-1 » bits destinée à indiquer la présence ou l'absence d'une copie d'un bloc ou d'une ligne mémoire dans les mémoires caches des multiprocesseurs des modules externes (51, 52, 53) et un bit de statut 5 Exclusif (87).

12. Serveur multiprocesseur à architecture multi-module selon la revendication 10 caractérisé le dispositif de commutation ou routeur (54) comporte des moyens de 10 gestion et/ou de filtrage des données et/ou requêtes en transit.

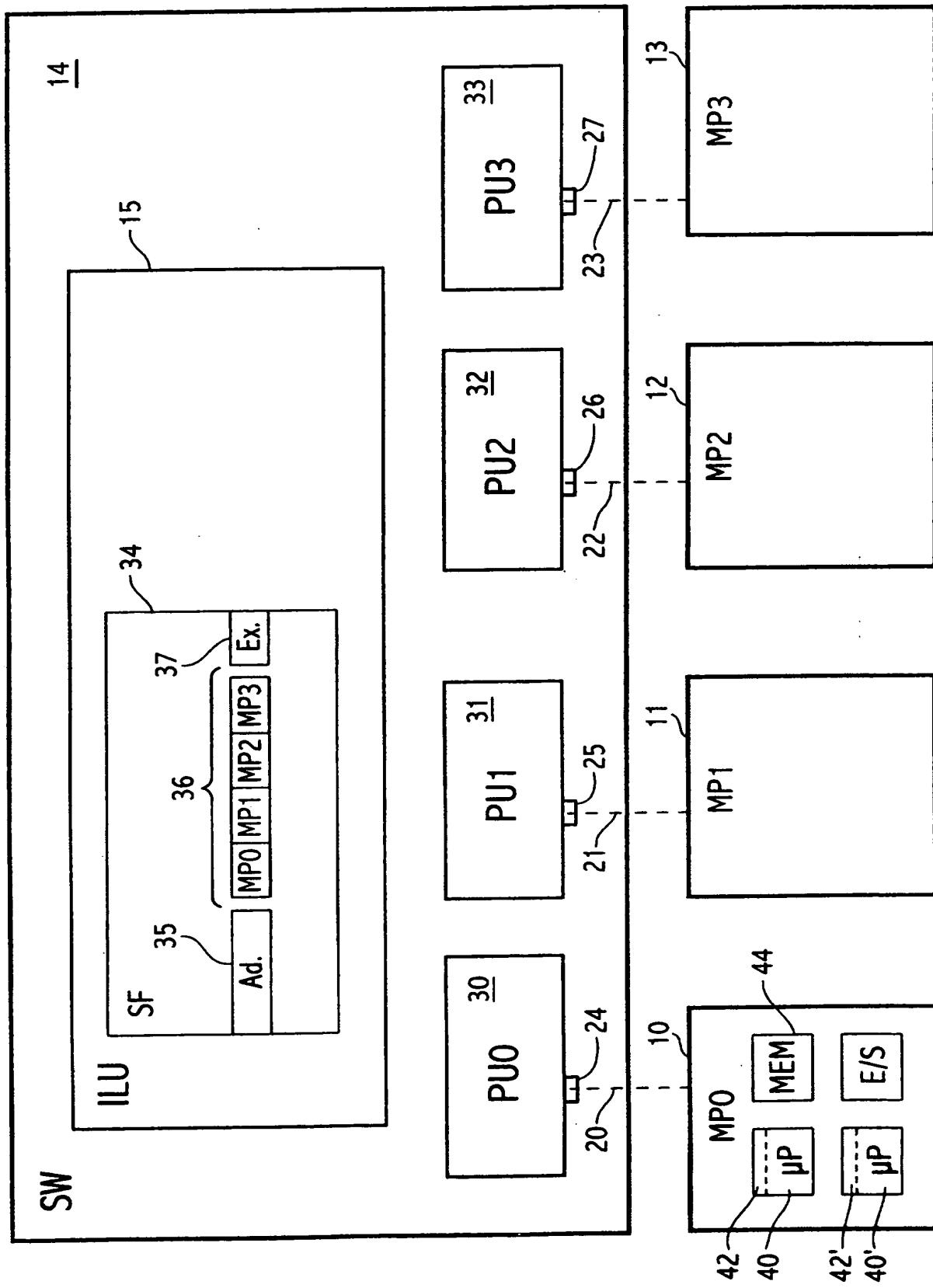


FIG.1

2 / 2

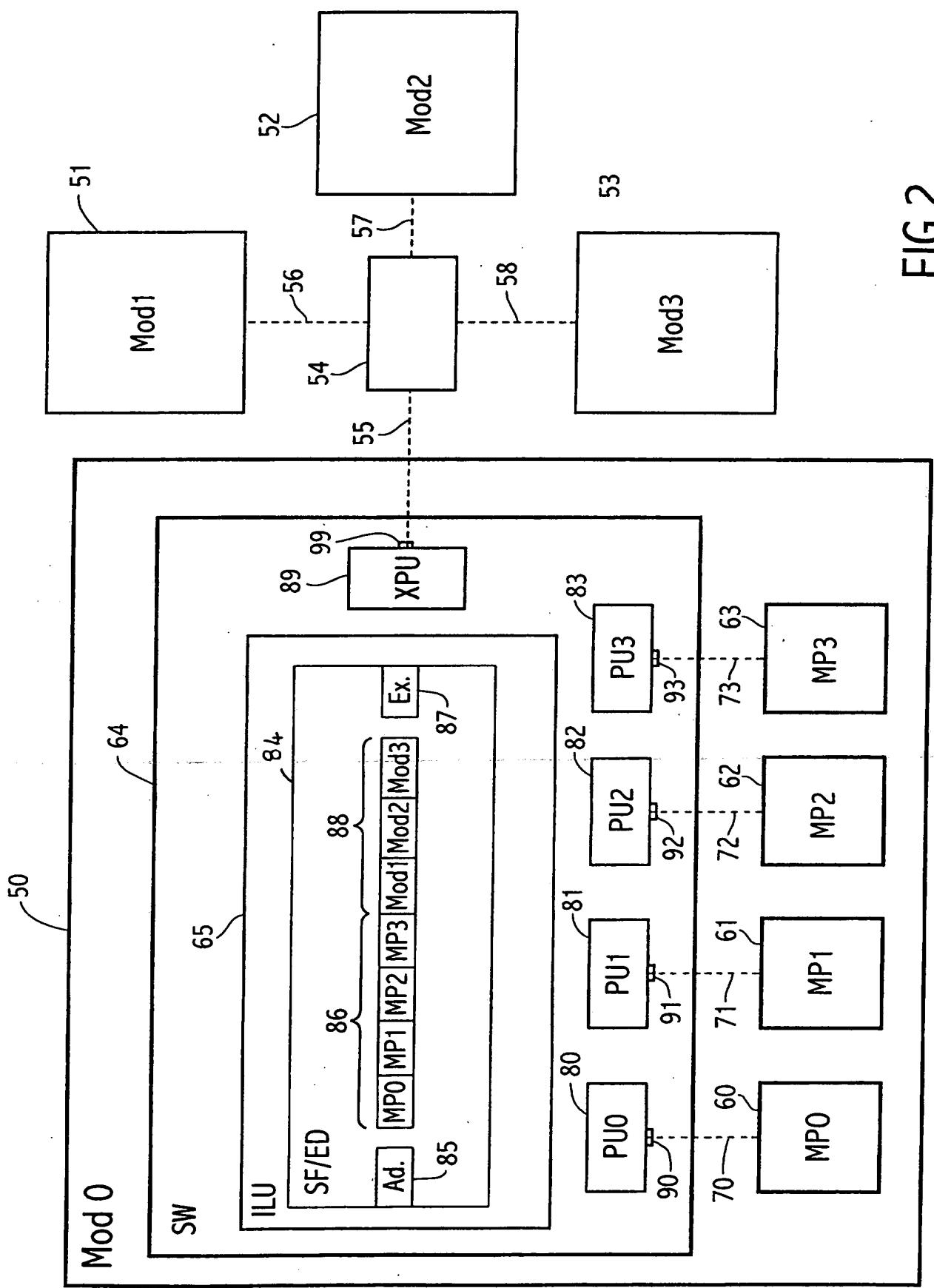


FIG.2

THIS PAGE BLANK (USPTO)

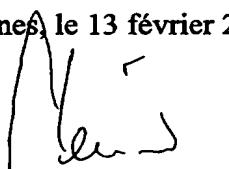
DÉPARTEMENT DES BREVETS

26 bis, rue de Saint Pétersbourg
75800 Paris Cedex 08
Téléphone : 01 53 04 53 04 Télécopie : 01 42 94 86 54

DÉSIGNATION D'INVENTEUR(S) Page N° .1 / .1
(Si le demandeur n'est pas l'inventeur ou l'unique inventeur)

Cet imprimé est à remplir lisiblement à l'encre noire

DB 113 W /260899

Vos références pour ce dossier (facultatif)	FR 3933 HD	
N° D'ENREGISTREMENT NATIONAL	0102083	
TITRE DE L'INVENTION (200 caractères ou espaces maximum)		
<p>Contrôleur de cohérence pour ensemble multiprocesseur, module et ensemble multiprocesseur à architecture multimodule intégrant un tel contrôleur.</p>		
LE(S) DEMANDEUR(S) :		
<p>BULL S.A. 68, route de Versailles 78430 LOUVECIENNES</p>		
DESIGNE(NT) EN TANT QU'INVENTEUR(S) : (Indiquez en haut à droite «Page N° 1/1» S'il y a plus de trois inventeurs, utilisez un formulaire identique et numérotez chaque page en indiquant le nombre total de pages).		
Nom Prénoms		Lesmanne Sylvie
Adresse	Rue	53, avenue Victor Hugo
	Code postal et ville	78340 Les Clayes sous Bois
Société d'appartenance (facultatif)		
Nom Prénoms		Koumou Pamphile
Adresse	Rue	35, avenue Jean Moulin
	Code postal et ville	78450 Villepreux
Société d'appartenance (facultatif)		
Nom Prénoms		Bernard Christian
Adresse	Rue	20, rue de la Commanderie
	Code postal et ville	78320 Le Mesnil Saint-Denis
Société d'appartenance (facultatif)		
DATE ET SIGNATURE(S) DU (DES) DEMANDEUR(S) OU DU MANDATAIRE (Nom et qualité du signataire)		Louveciennes, le 13 février 2001  Hervé DENIS (Mandataire Bull S.A.)

THIS PAGE BLANK (USPTO)